DOCUMENT RESUME

ED 462 765 EA 031 630

AUTHOR Beaton, Albert E.; Postlethwaite, T. Neville; Ross, Kenneth N.; Spearritt, Donald; Wolf, Richard M.

TITLE The Benefits and Limitations of International Educational Achievement Studies. Trends in Education.

INSTITUTION United Nations Educational, Scientific, and Cultural Organization, Paris (France). International Inst. for Educational Planning.; International Academy of Education, Brussels (Belgium).

ISBN ISBN-92-803-1188-3

PUB DATE 1999-09-00

NOTE 60p.

AVAILABLE FROM International Institute for Educational Planning/UNESCO, 7-9 rue Eugene-Delacroix, 75116 Paris, France ($10). Tel: +33(1) 45 03 77 00; Fax: +33(1) 40 72 83 66; Web site: http://www.unesco.org/iiep; e-mail: information@iiep.unesco.org. For full text: http://www.unesco.org/iiep/english/pubs/recent/rec11.htm.

PUB TYPE Reports - Evaluative (142)

EDRS PRICE EDRS Price MF01 Plus Postage. PC Not Available from EDRS.

DESCRIPTORS *Academic Achievement; Comparative Analysis; *Educational Development; Elementary Secondary Education; Foreign Countries; *International Studies

ABSTRACT

This booklet examines high-profile issues and major questions involved in international educational achievement studies. Some of these questions, which education activists and the public most often ask, are as follows: What benefits do international studies of achievement have for development of curricula in countries participating in this study? To what extent can the operation of a national system of curriculum (versus a state curriculum or school district curriculum or school-based curriculum) be assessed in an international study? If a comparison of subgroups of students within a country is to be undertaken, what are the subgroups that should be compared and why? Conceptualization and conduct of international studies are not an easy matter, but if they are well done, they can be of great benefit to national policymakers in education. Appendix 1 gives a brief description of the studies mentioned in this report. Appendix 2 alerts readers to the kinds of questions they should be asking when reading the reports of studies on international educational achievement. (Contains 47 references.) (RT)

Reproductions supplied by EDRS are the best that can be made from the original document.

ED 462 765

IIEP Observation programme

Trends in education

The benefits and limitations of international educational achievement studies

On behalf of the International Academy of Education

Albert E. Beaton, T. Neville Postlethwaite, Kenneth N. Ross, Donald Spearritt and Richard M. Wolf

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

J. Hall

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

EA031630

International Institute for Educational Planning

International Academy of Education

BEST COPY AVAILABLE

Trends in education

The benefits and limitations of international educational achievement studies

On behalf of the International Academy of Education
Albert E. Beaton, T. Neville Postlethwaite, Kenneth N. Ross, Donald Spearritt and Richard M. Wolf

International Institute for Educational Planning

3

International Academy of Education

The views and opinions expressed in this booklet are those of the authors and do not necessarily represent the views of UNESCO, the IIEP or IAE. The designations employed and the presentation of material throughout this review do not imply the expression of any opinion whatsoever on the part of UNESCO, the IIEP or the IAE concerning the legal status of any country, territory, city or area or its authorities, or concerning its frontiers or boundaries.

The publication costs of this study have been covered through a grant-in-aid offered by UNESCO and by voluntary contributions made by several Member States of UNESCO, the list of which will be found at the end of the volume.

Published by
International Institute for Educational Planning/UNESCO
7-9 rue Eugène-Delacroix, 75116 Paris, France
Typeset in France by Linéale Productions
Printed in IIEP's printshop

ISBN: 92-803-1188-3
© UNESCO September 1999

CONTENTS

		Pages
Preface		7
Explanatory note		9
Introduction		11
I	**Benefits and limitations of participating in an international achievement study**	13
	What benefits do international studies of achievement have for the development of curricula in participating countries?	13
	How can international studies relating education system characteristics to educational achievement levels contribute to the improvement of education systems?	14
	How can international studies of educational achievement contribute to the training of educational researchers in each country?	17
	What information is needed to assure decision-makers that the tests used in international studies are equally fair to all participating nations? Is there such a thing as a level playing field and how important is it?	18
	What can be done to ensure that international studies give priority to the major policy questions in all participating countries? How can decision-makers ensure that their particular concerns are addressed?	20
II	**Scope of research questions in an international educational achievement study**	22
	To what extent can the operation of a national system of curriculum (vs. a state curriculum or school district curriculum or school-based curriculum) be assessed in an international study?	22
	To what extent can variation in achievement among countries be related to different approaches to curriculum sequencing?	23
	To what extent can the operation of different types of examination systems be assessed in an international study?	24

To what extent can examining the variation among countries provide useful information on the effect of a factor that is invariant within countries? 25

To what extent does variation among school means at a given age level differ from country to country and what factors are associated with such variation? 26

To what extent do different countries change their relative position in school achievement levels over time and how important is this? 27

What is the relationship between achievement levels and economic productivity for a nation? 28

III Comparison within and between education systems in international achievement studies 30

How can achievement be compared across different age or grade groups and surveys several years apart? 30

If a country is involved with two agencies in international achievement studies and has its own national studies, what can it do to ensure that it can obtain a picture of as many grades as possible in several subject areas? 31

If a comparison of sub-groups of students within a country is to be undertaken, what are the sub-groups that should be compared and why? 32

IV General conclusion 34

Appendix 1 38

Appendix 2 41

References 50

General Bibliography 51

PREFACE

This publication was prepared by the International Academy of Education for inclusion in the IIEP Observation series. It is a publication designed for multiple audiences including education policy-makers, researchers, educationists, and the public. As the Academy's first publication in collaboration with the IIEP, it is appropriate that it be concerned with the very high-profile issue of international educational achievement studies. These studies are surrounded by complexity, misunderstanding, and questionable policy implications. This booklet examines the major questions and issues involved in international studies in a direct, concise and understandable manner.

The booklet includes the major questions that have arisen since the 1960s. All the major international studies are referenced, and examples used to demonstrate conclusions and explore trends. The authors present the evidence from several studies to answer the questions education activists and the public most often ask. Brevity is a particular virtue of this booklet in an area where technical discussions often obscure the overall trend or summary evidence.

Over 50 countries have participated in the studies utilized in this publication, including those in Africa, as well as the developed countries. The booklet ends with a brief section on the requirements of a good study.

The reader will conclude that we know quite a lot about the associations between national achievement and various national education characteristics and practices. But there remain large

domains where little is known, or the conclusions are tentative. This overview provides an agenda for future studies which the Academy hopes to influence. More international studies are needed, but they must focus on priorities, and be designed to yield sound results. We hope this booklet will be an important contribution to the entire field.

Michael Kirst
Professor, School of Education
Stanford University

EXPLANATORY NOTE

The IIEP's Medium-Term Plan contains provision for a range of 'Observation' activities, within which the Institute takes stock of emerging and future trends in the educational environment and then uses this information to revitalize and fine-tune its teaching, research, and training programmes. This form of dynamic feedback has drawn the Institute into involvement with studies that monitor international trends in educational planning and management, investigations that identify new educational needs emerging from social and economic transformations, and systematic self-reflection activities that seek to re-formulate the Institute's approach to international co-operation.

This programme of activities is indeed a challenging one, and for a small Institute like the IIEP it can only be addressed successfully by forming and sustaining productive partnerships with leading professionals around the world. To this end, the IIEP has been constantly broadening its international co-operation activities with a number of leading professional groups, institutes, universities, and agencies.

An excellent example of this has been the IIEP's establishment of strong research and training linkages with the Southern Africa Consortium for Monitoring Educational Quality. This powerful network of 15 ministries of education has been applying cross-national educational survey research techniques to the task of developing educational policies that will address concerns about the quality of education that have been expressed by ministries of education in Southern Africa. Another example has been the IIEP's formation of an alliance with city mayors, in developed and developing countries,

to design and present a conference on the problems of providing high-quality education in urban areas. This conference was subsequently expanded into an Internet-based global debate, or 'Netforum', where a much wider range of participants exchanged their experience and knowledge concerning the most innovative and successful methods for delivering education within large cities that are experiencing a multiplicity of financial difficulties.

This publication, entitled 'The Benefits and Limitations of International Educational Achievement Studies' and prepared by the International Academy of Education, also provides a very good example of the IIEP's continuing efforts to expand its programme of international co-operation and to keep its constituencies fully informed about the most recent developments in educational planning and management. Indeed, the Academy's contribution in this area comes at an opportune time. Governments in many countries have become increasingly interested in participating in international educational achievement studies, but at the same time they are seeking guidance as to whether the costs of participating in such studies can be justified in terms of their policy and pedagogical benefits. This significant document provides a well-researched, but very readable, analysis of benefits and limitations that can be used as a key resource for an informed debate concerning this question.

The IIEP would like to thank the International Academy of Education and the authors for their work on this document, and also to acknowledge the major contribution that they have made towards the IIEP's mission of improving the planning and management of education systems around the world.

Jacques Hallak
Assistant Director-General, UNESCO
Director, IIEP

INTRODUCTION

Although international studies of educational achievement began in the early 1960s, only in the 1990s have the results of such studies received recognition by governments and broad attention by the public because of widespread media coverage. An increasing number of countries have begun to participate in such studies, and in the most recently completed large-scale study of this type (the Third International Mathematics and Science Study (TIMSS) of the International Association for the Evaluation of Educational Achievement (IEA)), some 50 countries were involved. The IEA has conducted many studies in several subject areas for 40 years (see 'IEA studies' in the bibliography). The Educational Testing Service in the United States of America has also undertaken two large-scale international studies (1989, 1992) in its programme entitled the 'International Assessment of Educational Progress' (IAEP). The Organisation for Economic Co-operation and Development (OECD) has also commenced a project in this area by embarking upon a cyclical study of reading, mathematics and science for 15-year-olds that is known as the Performance Indicators of Student Achievement (PISA). The Southern Africa Consortium for Monitoring Educational Quality (SACMEQ) has been conducting and is continuing to undertake studies in several southern African countries. A brief description of these studies has been presented in *Appendix 1*. The main aim of all of these studies has been to collect and analyze valid cross-national information about student educational achievement.

The International Academy of Education has been asked to make a statement about the benefits and limitations of such studies for ministries of education. This document seeks to address this issue in

a manner that will be helpful to decision-makers in ministries, educational researchers, educationists in general, and the public. The Academy's main motivation for undertaking this task is that poorly-designed or poorly-executed studies provide little that is of use to ministries because they cannot be used in any meaningful way for policy analysis and development. The Academy's statement has been prepared as two inter-related components: a non-technical general response and a more technical appendix (see *Appendix 2*) that lists the 'Requirements of a good study' in a manner that indicates some of the points that readers should look for when reading the reports of studies in this area.

The statement has been structured around the questions that are most frequently posed with respect to studies in this area.

I BENEFITS AND LIMITATIONS OF PARTICIPATING IN AN INTERNATIONAL ACHIEVEMENT STUDY

What are the benefits and limitations to a nation or system of education in participating in an international educational achievement study? How can studies of cross-national achievement levels (and distributions of achievement) be used by educational administrators to improve their own system of education?

What benefits do international studies of achievement have for the development of curricula in participating countries?

Countries differ to varying degrees in the curricula they design for school students of a given grade level. These differences in curricula may also occur among states or regions in a particular country, or sometimes among schools within a region, or even among teacher-selected curricula within the same school. These differences may involve variations in the emphases given to different subject areas and/or variations in the emphases given to different topics or processes within the same subject area.

One of the major benefits of participation in international studies of educational achievement is the obligation it places on a country to subject its curriculum to close scrutiny. A curriculum in a particular subject area may become outdated in some respects and therefore fail to include topics or approaches of emerging importance. The close analysis and comparison of the curricula of several countries, which forms a prerequisite step in deciding which topics and test items should be included in a common test that is fair to all countries, impels a thorough review by each country of its own curriculum. In this way, ministries can be alerted to differences between their curriculum and that of other countries in terms of the emphasis

given to, and the content covered by, different subject areas. Such differences do not necessarily imply any deficiency because a country may well have good reasons for adopting a curriculum involving variations from the international pattern, for example reasons associated with the state of development of its education system.

The information provided on students' comparative performance on questions and sub-sections of a cross-national achievement test can be of direct benefit to a country in helping to determine whether to change the curricular emphases given to different subject areas and/or different topics. Changes in curricular emphases designed to improve achievement levels in one subject area, however, may involve a trade-off in other subject areas.

How can international studies relating education system characteristics to educational achievement levels contribute to the improvement of education systems?

International studies of educational achievement typically reveal substantial differences in the average level of student performance in different countries, whatever the subject area. Such information generates a question of more immediate interest to education systems. *Why* do some countries have higher levels of achievement than others? Most international studies collect information that allows for a detailed examination of factors likely to influence educational achievement.

Attention has mostly been focused on the following broad groupings of explanatory factors:

1 *Home background*: for example, the educational levels and occupations of parents, educational resources within the home

such as the number of books, daily newspaper, availability of dictionary, computer access, etc.

2 *School characteristics*: for example, type of school, school size, student/teacher ratio, etc.

3 *Teacher characteristics*: for example, the educational levels of teachers in terms of academic and pedagogical training, teaching experience, sex of teacher, etc.

4 *Teaching conditions and practices*: for example, class size, number of in-class instructional hours per subject area, specialist *v.* non-specialist teachers, amount of homework assigned, etc.

5 *Student motivation*: for example, liking for school, level of interest in subject areas and views on their importance.

The relationships among these factors and the educational achievement of students can be studied within individual countries, but international studies facilitate comparison of the *relative* importance of the factors in different countries. That is, international studies offer an important opportunity to replicate analyses across countries – which permits researchers to discover whether such factors affecting achievement are 'universally' important, or important for a group of countries, or important for a particular country. Such analyses usually commence with an examination of the strength of the (bivariate) relationships between particular background factors such as amount of homework assigned and achievement in a particular subject area. These bivariate relationships provide useful preliminary information about the likely importance of factors, but they need to be interpreted with caution because they may lead to simplistic and/or selective interpretations of relationships that fail to take account of the full picture – especially

15

the complex and multi-factorial nature of the education process. The fact is that many of these background factors are interrelated. The effect of any particular factor has to be teased out through multivariate analysis of the numerous bivariate relationships, using specialized techniques such as path analysis and hierarchical linear modelling. In the latter procedure, the data may be analyzed at multiple levels, which incorporate assessments of causal factors, for example, at the level of student performance, class performance, school performance, and country performance. That is, they attempt to provide a reply to the question: 'Other things (factors) being equal, what is the impact of this particular factor on educational achievement?'

A number of the factors listed above have been found in international studies to be associated with educational achievement across many countries after the influence of other factors has been taken into account. In assessing their potential value as sources of information for guiding the improvement of education systems, account has to be taken of the fact that some of these factors are more amenable to change through policy reform than others. To illustrate, sometimes rapid changes may be made in factors relating to curriculum and teaching conditions to improve an education system in the short term. For example, changes may be made in the amount of homework assigned, in the proportion of instructional time devoted to different subject areas, or in access to school library resources available for reading programmes. Longer-term change may involve, for example, changes in class size, the use of specialist teachers at secondary-school level, the adoption of teaching approaches designed to improve student interest and attitudes towards particular subject areas and, in some countries, increases in the level of teacher education and training. In some countries, increases in the number of years of schooling may be required to

achieve adequate levels of parental education and home literacy in the next generation. Although values instilled in the home and accepted by a particular society have a major influence on student achievement, changes in these values are much more difficult to bring about, and may not even be acceptable to the society in question.

How can international studies of educational achievement contribute to the training of educational researchers in each country?

Research studies involving a large number of countries with a diverse range of educational approaches are considerably more complex than those conducted within any one country. The diversity of educational environments associated with international studies demands close attention to methods of curriculum analysis, to procedures employed in the construction of measures of educational outcomes (knowledge, skills, and values) that are 'fair' to all participating countries, to issues of defining comparable populations and sampling procedures, and to the methodology required for data processing and statistical analysis of large numbers of variables from very large databases.

Training sessions in the use and interpretation of these procedures are routinely provided at the international level for researchers responsible for the individual country components of international studies. In turn, these researchers conduct training sessions for educational researchers in their own country. In this way, educational researchers in participating countries keep abreast of the latest developments in research techniques, while reports of the international studies make the techniques available to researchers in non-participating countries.

Researchers in specific subject areas of the school curriculum, such as reading, written composition, mathematics and science, can

17

also benefit from the wider perspectives taken in international studies and the detailed analyses which these studies have to make of the philosophy and practice of curriculum development and of the teaching approaches adopted in different countries.

What information is needed to assure decision-makers that the tests used in international studies are equally fair to all participating nations? Is there such a thing as a level playing field and how important is it?

In principle, a single set of test questions that is used to construct an international test cannot be perfectly appropriate for all students in all countries, in the sense that all students have had an equal opportunity to learn the subject matter that is tested. Curricula vary from country to country in terms of both content and sequencing and thus students always have different opportunities to learn. This variation in curricula, however, should not be seen as invalidating international studies because one of the important aims of these studies is to identify and document such differences. To achieve a fair basis for international comparisons, international tests should incorporate sub-groups of items that provide a valid assessment of most of the important components of each national curriculum. Two types of valid comparisons can then be made among countries. First, when two countries share one or more important common curriculum components, then the relevant subsets of test items may be used to make direct comparisons using the same items. Second, where it can be agreed that a single underlying trait (or dimension of knowledge) is being assessed by the test (or a subset of the test), then new technologies based on modern 'Item Response Theory' may be used to make valid comparisons of countries on groups of test items that 'overlap' but are not completely congruent.

Preparing a 'balanced' test that addresses the key (but not necessarily all) curriculum components of all countries begins with the inception of the study. Representatives of the participating countries should be actively involved in preparing the test specifications as well as in the development and selection of the actual test questions. Such involvement helps to ensure that the key curriculum components of the countries are covered and that the test questions provide a valid assessment of these components. When data are available from a pre-test, a number of statistical techniques are available to select test questions that might not be suitable for some countries, and these items can then be altered or eliminated. It is inevitable that the final test will contain some questions that are not completely suitable for all students in all countries. The important thing is that the test is sufficiently balanced to represent the major curriculum emphases in all countries.

When a test has been administered, it is useful to review its appropriateness for the participating countries. This can be achieved by asking participating countries to indicate whether or not the test items are appropriate for their students. The percentage of items deemed appropriate will always vary. However, provided that test items are used for comparisons in one of the two ways mentioned above, this will not affect the validity of comparisons. One of the most important research results arising from the IEA's TIMSS study was that the scoring of students on either items appropriate to their own curricula or items appropriate to the curricula in different countries did not substantially change a country's position in international standings.

At the lower grades in school, where school attendance is compulsory, the comparison of countries is relatively simple - even though there are many social and educational factors that may

contribute to student performance. An international test is an agreed-upon criterion that measures the net effect of all such factors. For example, countries may have different starting ages or retention policies, as well as curricula and pedagogical practices, that contribute to student performance and these must be considered when comparing countries' performance levels. At the end of secondary school, the differences among countries are more pronounced and obvious since systems vary substantially with respect to the number of grades required for graduation, apprenticeship programmes, drop-out rates, and the like. Such factors have to be taken into account when interpreting the differences in achievement levels. Perhaps the most important finding of an international assessment at this level is how school systems vary around the world. An international test provides a common method for assessing the output of the various systems when, as young adults, the students enter the workforce or higher levels of education.

What can be done to ensure that international studies give priority to the major policy questions in all participating countries? How can decision-makers ensure that their particular concerns are addressed?

There are two ways to ensure that major policy questions are addressed in international studies. First, international studies are by their nature co-operative ventures, and by joining a study at the very beginning, each country's representatives will have a chance to contribute to the planning of a study. Some studies (such as the Southern Africa Consortium for Monitoring Educational Quality (SACMEQ) studies listed in the bibliography) commence by developing policy questions common to all participating countries, and then plan studies to answer these questions (and only these questions). The SACMEQ approach ensures that policy questions that are of importance to all countries are addressed in the research design.

Second, individual countries may find that a policy issue that is of concern to them may not be of interest to other countries. If this occurs, it is possible for a country to add extra questions to the international instruments so as to provide measures required in order to address these issues. In the nomenclature of international studies, such nationally developed materials are often referred to as 'national options'. That is, an individual country is in a position to build a set of national options into an international study that will address its own policy concerns. The advantage of this strategy is that the country will be able to gather information focused on its own national issues, and at the same time obtain the benefits of participating in an international study which will relate its own national information to that of other countries.

21

II SCOPE OF RESEARCH QUESTIONS IN AN INTERNATIONAL EDUCATIONAL ACHIEVEMENT STUDY

What are the research questions that can be addressed in such studies that cannot be addressed in a national study or can be addressed only in a limited way?

To what extent can the operation of a national system of curriculum (vs. a state curriculum or school district curriculum or school-based curriculum) be assessed in an international study?

In international studies of educational achievement a distinction is often made between intended and implemented curricula. The intended curriculum of an education system is the planned curriculum and represents the expectations of what should be taught in a particular subject at a particular grade level. Persons responsible for the intended or planned curriculum at a national level should be in a position to review individual test items to determine whether they are testing what is included in the national curriculum. By engaging in such a review, it is possible to identify those parts of an international test that are a part of the national curriculum in a particular subject at a particular grade level. It is then possible to develop a test score based on that nation's curriculum. Furthermore, it is possible to compare a country's performance on that subset of questions and items from an international test with that of other countries on that same subset of questions and items.

The point here is that a country can obtain scores focused on its own curriculum for all other participating countries. This offers an important avenue for making what many educationists believe to be 'fair comparisons'. Research evidence shows that while these fair comparisons are important, there are major similarities in curricula

across countries. That is, the central elements of a national curriculum in a particular school subject are often quite similar across countries. What does differ, often substantially among countries, is the timing of the teaching of particular topics and skills. Differences of as much as two or three years in the introduction of particular topics and skills are not uncommon.

Some countries have a centralized and uniform national curriculum for some or all subjects at various grade levels. Other countries have regional or state-level curricula, and still others have flexible curriculum frameworks that permit school-based control over curriculum details. International studies allow comparisons to be made both at a national and sub-national level. Some people have held the view that a centralized and uniform national curriculum will lead to higher levels of achievement compared with more flexible arrangements. However, the evidence from the IEA TIMSS study has indicated that at Grade 7 and 8 levels there was no relationship between whether a country had a national curriculum or not and how well its students performed in science and mathematics.

To what extent can variation in achievement among countries be related to different approaches to curriculum sequencing?

Different countries adopt different approaches to curriculum sequencing, particularly in the area of science. In contrast to within-country studies, international studies offer the possibility of comparing the achievement levels of students following these different approaches. Students in Grades 8 or 9, for instance, study courses in general science in many countries, but separate courses in biology, chemistry, physics and sometimes Earth science in other countries. Data from the IEA's Second International Science Study (SISS) suggest that students in countries where the separate courses

are compulsorily studied gain higher achievement scores, though this may partly reflect differences in opportunity to learn the content covered in the science achievement test (Wiley; Schmidt; Wolfe, 1992). International surveys also suggest that at the lower and middle-secondary school levels, a curriculum requiring the simultaneous study of all of these branches of science leads to a greater growth in science achievement than one based on consecutive study of the different branches of science (Keeves; Soydhurum, 1992).

More detailed information on cross-national variation in curriculum sequencing has been obtained in studies of 'topic trace mapping' in the TIMSS study. These studies noted, for instance, the year level in which topics were introduced in different countries and the number of years over which those topics were studied. Country variations in the inclusion and sequencing of topics in the curriculum, as distinct from offering branches of courses simultaneously or consecutively over a number of years, however, do not appear to have any appreciable effect on cross-national achievement levels. Test-curriculum matching analyses in TIMSS showed that the pattern of results for the different countries was much the same, regardless of whether a country was compared with other countries on the basis of its own curriculum, or in terms of the curriculum followed by each of the other countries.

To what extent can the operation of different types of examination systems be assessed in an international study?

There has been a long-standing debate in many countries about the relative merits of internal (school-based) examinations that are set and graded within schools, and external examinations conducted throughout a country or province by state authorities or examining boards on prescribed syllabuses. Consequently there has been much

interest in data from international studies about the comparative achievement levels of students in countries adopting one or other of these two types of examination system.

Data from the IEA TIMSS study have shown that 13-year-old students in most countries with a curriculum-based external examination in mathematics at the end of the final year in secondary school have tended to do better in mathematics tests than their counterparts from countries with no such examination, and likewise with science students (Bishop, 1998). Presumably, the effect of an external examination at this point is not limited to achievement in the final year of secondary school, but flows down to earlier years in the school system.

While the effects of different types of examination system can be estimated in this way, some caution should be exercised in the interpretation of results. The data show an association between levels of achievement and the presence or absence of curriculum-based external examinations, but not a causal relationship. Other possible explanations of such associations, such as the use of specialist versus non-specialist teachers, for example, have to be ruled out either by statistical controls or inspection of other likely variables of relevance. The above results have been shown to apply when the most likely alternative explanatory variables have been taken into account.

To what extent can examining the variation among countries provide useful information on the effect of a factor that is invariant within countries?

One of the important motivations for undertaking international studies of educational achievement is to assess the effects of factors that are invariant within countries but vary across countries. A notable example of this is the age of starting school. In the first IEA

International Study of Mathematics Achievement (Husén, 1967), some participating countries commenced formal schooling at age five, others at age six, and still others at age seven. The researchers in that study were interested in the consequences of different starting ages. This question was investigated by comparing the achievement test performance of students in the three starting age groups at age 13. The results showed that there was no difference between starting at age five or six, but that students in countries that started school at age seven scored slightly lower at age 13. In a similar way, the effect of other factors that are invariant within countries but vary among countries can be assessed.

To what extent does variation among school means at a given age level differ from country to country and what factors are associated with such variation?

Some of the variation in student achievement test scores in a country is due to differences in the performance of students in the same class or same school, while some of the variation is due to differences in the average performance of students in different classes or different schools. International studies have shown that school means differ much more widely in some countries than in others.

In a 1981-86 international study, for instance, the average science achievement level was much the same for all classes, both at Grades 4 to 5 and Grades 8 to 9 in Finland, Japan, and Sweden. In contrast, school means differed considerably for one or both of the grade groupings in Italy, the Netherlands, the Philippines, Ghana, and Nigeria (Postlethwaite; Wiley, 1992). In Italy and the Netherlands, the large differences among school means could be attributed to the 'streaming' of students into different types of schools at Grades 8/9.

In the other three countries, the large differences among school means probably reflected urban-rural differences in teacher qualifications, school resources and equipment, and the degree of parental interest in education.

Information of this kind, when studied for countries at a similar stage of development, can be useful to educational administrators in reviewing the structure of their own systems - and the issue of equity among schools.

To what extent do different countries change their relative position in school achievement levels over time and how important is this?

Countries will often change their relative position in relation to other countries over time. Small changes occur frequently and may simply be due to chance. These changes should not be taken seriously. It is important, however, that groups conducting such studies determine whether the changes that occur are due to chance or not and, if not, how meaningful they are. (Since there are large numbers of students involved in such studies, almost any change is apt to be statistically significant. This does not mean that the change is meaningful.) It is also important to determine whether the changes are due to changes in the education systems or to changes in assessment methodology as, for example, when population definitions are changed or when the countries that participate are quite different.

In some cases, countries will show substantial change in their relative position over time. Such changes may be significant and meaningful. It then becomes a challenge to the researchers and the individual country to determine the likely causes of such change. One example of a substantial change occurred from the first to the third international mathematics studies (Husén, 1967, and Beaton et

al., 1996). In the first international mathematics study, USA 13-year-olds scored eleventh out of twelve participating countries, while in the third international mathematics study, USA students scored twenty-sixth out of forty countries, a notable improvement. In fact, 20 countries scored significantly higher than the USA while seven countries scored significantly below the USA and 12 countries scored at essentially the same level as the USA. Such a change should be carefully examined to establish reasonable explanations for the difference, while keeping in mind that a change in the relative position of one country may be due to changes in the other participating countries as well as changes in the performance of the particular country's students.

What is the relationship between achievement levels and economic productivity for a nation?

There is very little firm evidence to support the widely-held view that there is a strong and direct causal connection between mean student test scores for nations and their economic competitiveness (Levin, 1998). A simple bivariate association between these two factors using means aggregated to the national level is yet to be demonstrated reliably - and even if it could be demonstrated, then this would not identify which factor was 'cause' and which was 'effect'. For example, a weakened economy may reduce the quality of the conditions of schooling and thus precipitate lower educational outcomes. Or, it may be the case that a complex interaction of political and social factors will trigger impacts on both educational outcomes and economic performance (Goldstein, 1994). There are many factors associated with higher national productivity, not least of which are sound governance, skilful management of industry and commerce, the alignment of established cultures with the requirements of a modern workforce, and the vagaries and volatility

of international markets. At the same time, it is common sense that national economic success can be affected by effective decision-making at all levels of governance and commerce, which in turn demands appropriate knowledge, skills and attitudes among those involved in the enterprise.

The studies reviewed in this report can provide a great deal of assistance in identifying factors within an education system at one point in time that are associated with higher levels of achievement. Some of these factors are 'malleable' (in that they are amenable to change by government policy) and others are not. For example, ministries of education can make decisions about the resource levels (material and human) that can lead to improved learning outcomes. However, governments have negligible short-term influence over improving the educational and economic conditions of students' homes. Of those that are malleable, some can be best managed by ministries of education and others by lower levels within the school system. The SACMEQ studies mentioned at the end of this report each have conclusions listing policy suggestions, categorized by the time and cost to implement them, some of which are aimed at the ministries of education and linked to responsible persons or groups within ministries of education and some of which are aimed at the regional levels within countries or at schools.

29

III COMPARISON WITHIN AND BETWEEN EDUCATION SYSTEMS IN INTERNATIONAL ACHIEVEMENT STUDIES

What meaningful comparisons can be made within and between different systems of education in any one study (or between studies)?

How can achievement be compared across different age or grade groups and surveys several years apart?

The issue of how one can compare achievement across different age or grade groups and surveys several years apart is an important one for countries that are considering whether to participate in an international study of achievement. Governments are interested in knowing if the level of student performance is improving, holding steady, or declining. Advances in the technology of testing over the past several decades make these comparisons possible without having to administer exactly the same test to students in the groups to be compared. What is needed to make such comparisons is a small set of well-calibrated common test questions (embedded within any pair of the different tests) whose substantive and statistical properties (in terms of relative difficulties) are known and are stable. To achieve this, extensive statistical and psychometric analysis of test score data is needed in the phase of test development, and steps need to be taken to keep the common embedded test questions secure. Such work permits the kinds of comparisons that countries want to make without being restricted to administering exactly the same test on different occasions.

If a country is involved with two agencies in international achievement studies and has its own national studies, what can it do to ensure that it can obtain a picture of as many grades as possible in several subject areas?

National assessment studies are concerned with measuring achievement in particular subject areas within a nation at particular grade levels. This can focus on assessing achievement levels at one point in time, or in assessing change over time. In national assessment studies, one major focus is on the differences among sub-groups of students or schools or regions within countries. The sub-groups used in these analyses are often nation-specific. The major thrust of international studies is to compare achievement among nations. There is often little focus on differences among sub-groups of students or schools within countries. Nevertheless, the data from international studies can often be used to conduct important national analyses. Both national and international assessments are costly, especially when the opportunity costs of student and teacher time are included. Although international studies are uniquely appropriate to answer questions about how a country compares with other participating countries, it is important for administrators to ensure that they gain maximum benefits for their involvement. The surest way is for a country to get involved in an international study at the beginning so that it has an important part in shaping the study's design - including the grades or ages that are sampled and the subject matter and topics being assessed. No single country can be permitted to dominate the development of an international study, and so no single country is likely to achieve all that it wants from an international study. A country must be selective in choosing the studies in which it will participate.

In some cases, it will be possible to augment an international study with probes into national issues through additional test booklets or questionnaires that address national issues. If national and

31

international concerns overlap sufficiently, it may be possible to drop the national testing completely and rely on the international study for intra-country as well as inter-country analyses. However, in many cases national and international studies are designed for different purposes and investigate different populations and subject matter, and therefore a choice must be made on the basis of potential benefits.

It is often tempting to link national and international tests in order to infer how students who took a national test would have performed if they had taken the international test. This is especially the case where ministries of education wish to know how many students in one country would exceed some performance level reached by students in another country. Formally, this cannot be done precisely unless the two tests were built to precisely the same specifications, which is both difficult and unlikely. If the two tests are reasonably similar in that they share some items in common that have certain statistical properties, useful estimates can often be made through statistical linking, although the estimates may not be as accurate as one might wish. In doing linking studies, it is important to calculate the correct estimates of sampling error in order to avoid making uninformed policy decisions.

If a comparison of sub-groups of students within a country is to be undertaken, what are the sub-groups that should be compared and why?

In countries consisting of a federation of states or provinces, comparisons of achievement among states can be useful to educational administrators and governments in reviewing their education systems and determining how resources might be best employed to strengthen state systems.

In general, equity of educational provision should be the criterion employed in selecting sub-groups of students for comparison within

a country. In most cases, the administrators will wish to compare students who, on the basis of previous experience, are known to be members of groups of students that (a) differ in terms of educational opportunities, and/or (b) come under the administrative control of a single sub-office of the Ministry. Gender is often used as a key basis of comparison in order to check that girls and boys have equal educational opportunities. Gender comparisons should include not only achievement levels but also such factors as participation rates in schooling and attitudes towards subjects such as mathematics.

Comparisons between the achievement and participation rates of students from urban and rural schools or districts within a country can also be of value to administrators in distributing resources, particularly in developing countries where school location is known to be strongly linked with educational outcomes. This information can guide decisions aimed at achieving more equitable educational provision. For the same reason, information about the performance of minority language or cultural groups within a country in comparison with mainstream groups provides a useful guide to the type and extent of resources required for their education.

In countries with substantial proportions of students in non-government schools, government school versus non-government school comparisons may be useful, both as a guide to governments on school system standards, and as a guide to parents. Such comparisons, however, need to take account of differences in the socio-economic background of students.

Comparisons of student performance by socio-economic levels may be of general interest, but unlike the sub-group comparisons noted above, socio-economic level is less readily amenable to change through resource allocation by educational administrators or governments.

IV GENERAL CONCLUSION

Many critics of international studies of educational achievement have raised problems such as 'is like being compared with like?', 'is it really possible to make a fair test for students at, say, Grade 8 in many countries given that curricula vary?', 'are such studies useful?', and so on. The International Academy of Education was asked to write a short document pointing to the benefits and limitations of such studies.

As can be seen from what has been written above, the members of the Academy's Task Force have stated that there are many benefits to such studies *on condition that* the studies have been well conceptualized and conducted. To this end, the Task Force has prepared a short Appendix to this document that alerts readers to the kinds of questions they should be asking when reading the reports of such studies. It is clear that the conceptualization and conduct of international studies of educational achievement are not an easy matter, but that if they are well done, then they can be of great benefit to national policy-makers in education.

Even when studies are scientifically sound, there is need to exercise care when interpreting the results. For example, readers must be careful not to assume that a relationship established at one level of data aggregation (say, among countries) is the same at another level of analysis (say, among students within countries). Readers making assumptions of this kind are in danger of drawing false conclusions because of the so-called 'ecological fallacy'. This fallacy often resulted in erroneous conclusions being drawn by ministries and the media from the initial TIMSS reports (Ross, 1997).

It is always desirable for a country to keep its own curriculum under review and it is a salutary experience to compare a country's curriculum in a subject area with that of other countries. Many pointers for the judicious improvement of national curricula can be gleaned from such analyses. Examples were presented above of the kinds of information that can be useful to policy-makers about the structure of the curriculum and its relationship to levels of achievement.

It is possible to identify those aspects of schooling that have an effect on achievement but this requires special, and often complex, relational analyses. It is dangerous to try to derive better practices simply from examining nationally aggregated data for a few countries. In most large-scale studies it has not been possible to identify teacher methods and behaviours that have a strong effect on achievement – when these have been based on cross-sectional studies or longitudinal studies of one year's duration. The potential influence of these variables is usually better examined in replicated small experiments.

It was indicated how tests can be developed that allow achievement to be compared both across countries and across designated sub-groups within countries. It was interesting to note in the TIMSS study that the rank order of countries remained substantially the same even when the comparisons were made on the curricula of different countries.

Even though countries are participating in an international study they may add special test items and questionnaire questions that are of national concern, thus using one data collection for national and international purposes. Countries wishing to do this have to be prepared to make the required effort. The SACMEQ studies, for instance, were focused on policy issues that were identified some time before the commencement of the research design. Each SACMEQ

national report concludes with a set of policy suggestions focused on these policy issues. Ministries of education appreciate this kind of study because they can use the results to make informed policy decisions.

Some countries have been involved in more than one international study, as well as having their own national studies, and they often ask about comparing the results of the different studies. If this is to be done, then the different studies must have sufficient common test items. Unfortunately, there has been very little co-ordination in the design and conduct of different international and national studies, so that it is often not possible to make such comparisons. But, with some co-operation among the projects, this could be achieved.

It is often assumed that achievement levels are associated with economic productivity. There is little hard causal evidence on this point, although it would seem to be common sense that the more students are educated the more they should be able to adapt to new job requirements that changes in a labour market might demand. The types of studies conducted by organizations such as IEA are focused on the variables that might improve achievement in a current system of education and not variables that might improve economic productivity, which would clearly be affected by many non-school factors.

National research personnel can benefit enormously by being involved in well-conducted international projects. The scale of these projects requires the participation of experts in the many different skills needed to execute studies of this kind.

In short, the members of the Task Force were convinced that these kinds of studies are worthwhile. But, they do require effort on the part of the participating countries, a great deal of expertise on

the part of the researchers, and great care in interpretation by the researchers and policy-makers. Recommendations for policy changes in a country need to take account of not only the results of the international analyses, but of the educational and cultural context in which that country operates.

37

APPENDIX 1

Brief description of studies mentioned in this report

- IAEP

The International Assessment of Educational Progress (IAEP) consisted of two international assessments of mathematics and science that were conducted in 1988 and 1990-91. Seven countries participated in the first study and 20 in the second; 13-year-old students were tested. These studies were conducted under the auspices of the Educational Testing Service (ETS) in Princeton, NJ.

- IEA

The International Association for the Evaluation of Educational Achievement (IEA) is an international non-governmental organization whose members are research centres and ministries of education. Since 1959, it has conducted several assessments of reading, mathematics, and science as well as assessments of other subjects (see *General Bibliography*). Its studies are designed to provide policy-makers, educators, researchers, and the general public with information about educational achievement and its learning context. The countries presently in IEA are: Australia, Austria, Belgium (Flemish), Belgium (French), Botswana, Brazil, Bulgaria, Canada, Chile, China, Chinese Taipei, Cyprus, Czech Republic, Denmark, England, Finland, France, Germany, Greece, Hong Kong (SAR of China), Hungary, Iceland, Indonesia, Iran, Ireland, Israel, Italy, Japan, Kenya, Korea, Kuwait, Latvia, Lithuania, Luxembourg, Mexico, Netherlands, New Zealand, Nigeria, Norway, Philippines, Portugal, Romania, Russian Federation, Scotland, Singapore, Slovak Republic, Slovenia, South Africa, Spain, Sweden, Switzerland, Thailand, Turkey, and the USA.

PISA

PISA is the OECD 'Programme for International Student Assessment'. PISA is a collaborative process. It brings together scientific expertise from the participating countries and is steered jointly by their governments, through the OECD, on the basis of shared, policy-driven interests. The target population is all 15-year-olds in each of the participating countries. It aims to assess how far students approaching the end of compulsory schooling have acquired some of the knowledge and skills that are essential for full participation in society. Assessments will take place every three years. Three 'domains' form the core of each cycle: reading literacy, mathematical literacy, and scientific literacy. The first testing will take place in the year 2000, the second in 2003, and the third in 2006. The countries involved are: Australia, Austria, Belgium, Brazil, Canada, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Latvia, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Russian Federation, Spain, Sweden, Switzerland, United Kingdom, and the USA.

SACMEQ

The Southern Africa Consortium for Monitoring Educational Quality (SACMEQ) is a network of ministries of education in the Southern Africa sub-region that has the status of a non-governmental organization. Its primary aim is to provide training programmes that are carefully designed to build the capacity of educational planners in ministries of education to undertake large-scale educational policy research. SACMEQ training programmes are presented in a 'learning-by-doing' mode in which educational planners are involved in cross-national co-operative educational policy research projects that generate information which can be used by decision-makers to plan

the quality of education. Its first major study was undertaken in seven countries during 1995 and was focused on a baseline assessment of the conditions of schooling in Southern Africa and an evaluation of the reading literacy levels of Grade 6 students. SACMEQ's second major study commenced in 1999 and will be concerned with an evaluation of Grade 6 mathematics and reading achievement in Botswana, Kenya, Lesotho, Malawi, Mauritius, Mozambique, Namibia, Seychelles, South Africa, Swaziland, Tanzania (Mainland), Tanzania (Zanzibar), Uganda, Zambia, and Zimbabwe.

TIMSS

The Third International Mathematics and Science Study (TIMSS) is the largest international assessment ever administered. In 1994 and 1995, mathematics and science achievement were tested for over 500,000 students in the more than 40 participating countries. In many countries, students in Grades 3, 4, 7, 8, and 12 were tested. TIMSS was run under the auspices of the IEA. The countries that participated and participate in TIMSS are much the same as those mentioned as members of IEA above. The TIMSS international results, including technical reports, are available on the web at www.csteep.bc.edu/timss.

APPENDIX 2

Requirements of a good study

This appendix mentions several important matters that must be taken into consideration in order to judge the quality of large-scale international studies of educational achievement. It would be possible to write a very detailed book on all of the technical standards to be respected when conducting such studies. What follows, therefore, is a selection of the more important aspects.

▒ Aims of the study

Are the aims clearly stated? Is evidence presented in the reports to show that the research questions that have been developed address important policy and theory-oriented issues in many of the participating systems? (If this is not the case, then there is a danger that the research issues are the favourite topics of the researchers rather than those of the practitioners.) Is there evidence to show that the design of the study was specifically developed to allow the policy and theory-oriented issues to be answered? (In the SACMEQ studies, for example, great effort is invested in the identification of the policy issues common to many systems of education. The research questions are developed to answer the policy questions and then 'blank' or 'dummy' tables are developed to answer these research questions at the onset of the study.)

The reader needs to be assured that these steps were taken.

▒ Design of the study

Does the report of the study indicate how the design was developed in order to answer the questions posed? Do the

instruments that were developed cover all of the questions? Does the study design offer trade-offs between the complexity required to respond to specific questions and the need for simplicity, timeliness, and cost-effectiveness? Is the study design focused on the issues in question? Does it avoid unnecessary questions in the instruments that may relate only to peripheral research interests? If the study purports to help educational policy-makers and administrators, was the sampling conducted in such a way that good estimates of the variables in question could be produced for the national-level statistics, as well as for the administrators at the state and regional levels in cases where estimates for sub-national entities are desired?

▩ Target population definition

Where comparisons are being made across countries, is like being compared with like? For example, if students in a specific grade group are being compared for their achievement, are all of the students in the grade included in the target population or have some countries missed some out? It is usual to have some students 'excluded' either because they constitute a very small number and it would be exorbitantly expensive to collect data from them (for example, in very isolated areas) or because they are in special education schools (for example, for the blind or deaf). These students are normally referred to as the 'excluded' population. It is normal to keep these 'excluded' down to less than 5 percent of all students in the 'desired' target population. What is not desirable is to have 2 percent excluded in some countries and 14 percent in others. Are instruments in place to quantify the extent of school and student-level exclusions and to assess the impact of these exclusions on comparisons of means and distributions of the criterion variables?

The same argument applies when age groups are being compared. One argument for using age groups rather than grade groups is to discover the achievement of the students born between certain dates (for example, a calendar year). This approach seeks to examine how systems of education have coped with the education of an age cohort. Where systems have high rates of grade repeating, it is possible to have students of, say, age 13 or 14 spread across several grades. Some systems will argue that the tests are too difficult for those students who are three grades behind the others and these should therefore be 'excluded'. In this case, either the tests do not have enough 'bottom' to them (in which case it can be argued that the tests are not appropriate for all of the students), or the students should be awarded zero or chance scores. One way of dealing with this problem is again to apply the rule that not more than 5 percent of the students should be 'excluded'.

When the reader has examined how the above problems have been dealt with by the researchers, then he/she can decide on whether like is being compared with like.

▩ Sampling

The object of sampling is to sample the defined target population in such a way that the population estimates derived from the samples have a sampling error that is acceptable with respect to the policy decisions that will be based on the results. In recent years, many of the major international studies have adopted the standard of having sample designs that have the same or better sampling precision as a simple random sample of 400 students for educational outcome measures. This level of sampling precision provides sampling errors for results on test items (percentage correct) of no more than 2.5 percent for one standard error and no more than 5 percent for two standard errors. This means, for example, that for a population

estimate of 50 percent, then one can be sure, 19 times out of 20, that the true value of the 50 percent lies between 45 and 55 percent.

In some studies where direct international comparisons are less important, a lower level of precision may be acceptable – this is often the case where a broad analysis of the curriculum is required. Such studies often aim at discovering 'well-achieved', 'averagely achieved' and 'poorly achieved' objectives. In this case, it might be sufficient to say that 'well achieved' means about 80 percent of the students have mastered a particular objective or curriculum unit, 'averagely achieved' means about 50 percent, and 'poorly achieved' means about 30 percent'. In this case, two standard errors for results on test items (percentage correct) could have a magnitude of 10 percent – and this outcome would be good enough for the purposes of the study.

The question for the reader then is 'Was the sampling conducted in such a way as to yield standard errors of sampling that are acceptable for the purposes of the study?' It is usually the case that researchers who are knowledgeable in the area of sampling provide a detailed description of the steps of sampling and the correct sampling errors – which have been calculated to acknowledge any complexities in the sample design. If this information has not been provided, then there is a distinct possibility that the samples are suspect. It is also usual for the standard errors of sampling to be presented in the tables of results. If they are not there, then the reader should be wary.

Instrument construction

It is usual to have tests, questionnaires, and attitude scales as the measuring instruments in studies of this kind. Some studies may include observation schedules and interviews, but these are rare and will not be treated here.

The *test* instruments must cover the intended curriculum of the participating countries. This normally involves a two-stage process: first a content analysis of the curricula (via curriculum guides, textbooks, examinations, and what teachers say they teach) in the various countries; second, an arrival - on the basis of the first step - at the international blueprint for the tests. While many of the curricular objectives will be common across countries, some countries will have some objectives that are common to only a subset of countries. Mathematics and science are the key subject areas in which there is considerable curricular variation among countries. There is much less variation in subjects such as reading and foreign languages. There must, however, be agreement on the international blueprint and this must cover the bulk of the curricula in all countries if it is the intention of the study to focus on the common contents of national curricula.

Test items must be written to cover all cells having objectives in the blueprint. The item formats must be agreed and justified. The items must be trial-tested and analyzed. Where multiple-choice items are used then the distractors must be plausible not only in terms of content but also in their distracting power. Free-response questions requiring students to construct answers should be pre-tested to ensure that they will yield a range of responses that can be reliably scored. Where scaling is being used there must be agreement on the substantive meaning of the scale in terms of student performance on specified tasks at specified points of the scale. There must be agreement on the appropriateness of the items and the tests must be shown to be reliable. Where there is an attempt to measure change over time, say from the last survey to the current one, then there must be sufficient common items between the two points in time to allow change to be reliably measured. Finally, items should be tested for item bias in each and every country. The psychometric properties

45

of the test items should be similar over a sufficiently large number of countries. Where overlapping tests have to be used it must be shown at the trial stage that the common items used to allow calibration on to the same scale fulfil their purpose.

In some instances, 'hands-on' performance assessment tasks may be deemed necessary to cover the full range of objectives in a subject area. The design of such tasks should take into account the (usually) limited amount of time available for testing, the need to make use of equipment which is simple and available in multiple copies and not beyond the resources of participating countries, and the need to yield responses that can be reliably graded across countries.

The *questionnaire* instruments must include questions to cover all of the indicators needed to answer the policy issues raised at the onset of the study. Several of the indicators needed will be what are normally called 'derived variables' – that are constructed from the information obtained from several questions. The questions must be written in a simple language easily understood by all of the students (able and less able) who have to answer them. All questions must then be trial-tested and analyses undertaken to ensure that the questions are providing accurate and reliable information for the indicators and 'derived variables'.

The *attitude* instruments, sometimes a part of the questionnaires, measure selected attitudinal dimensions. The dimensions must be described. Attitude items are normally collected through special small studies from the target population members. They too are trial-tested and analyses undertaken. Very often about three times as many items are needed for trial testing as for the final attitude scale measure. The final scale must be shown to be reliable for the purposes for which it is intended.

The *translation* of instruments is not an easy business. Procedures must be put in place that ensure the cultural appropriateness and linguistic and psychometric equivalence of the instruments in participating countries. Normally there is a set of procedures that are followed for translation. Quite often two forward translators are used (from the language of the international test into the language of the national test) and where there are discrepancies, these must be ironed out. The final national language version is then back-translated into the language of the international test or the translated versions are checked by a central team having experience in all of the languages being used. It is through these checks as well as through the technical type of data analyses that poor translations are identified. Again the translation process is usually described in one of the publications in some detail.

Some of the points raised above are very technical but if they have been covered they will be explained in the report. What is important is for the reader to make sure that they have been discussed. If they have not been described in one of the study's reports, then there is ground for scepticism.

Data collection

The data-collection stage in each of the countries is crucial. The object of the data collection is to test all respondents selected in the sample and to have them complete every question in the questionnaires and all test items that they are able to answer. Normally, a manual is written for the persons in charge of the data collection at the national level in each country. This manual is required so as to ensure that the data-collection procedures proceed in a manner that will provide valid data under conditions that are uniform at each data-collection site.

47

The data-collection manual should cover every possible detail that must be taken into account when conducting the data collection. This involves 'school forms' and 'student forms' to ensure that the correct schools are selected, the correct students are tested (and not others), and the correct teachers are selected (where questionnaires or tests are being administered to teachers). A second manual is usually prepared for the data collectors and details everything to be done within each selected school. A third manual spells out (a) what each test administrator has to do and say during the actual testing sessions, (b) the procedures and timing for the administration of the instruments, and (c) how to parcel up the instruments and return them to a central point. There should be very few, if any, missing schools and very few missing students in the data collection. Again, the authors of the reports should report the percentage of missing schools and missing students. It is often said that not more than 10 percent of schools should be missing from the sample and not more than 20 percent of the students. However, since there are no completely valid procedures for dealing with missing data, these figures should be taken as absolute maximum levels.

Data recording, data cleaning, and the calculation of sampling weights

The data are usually recorded on computers at the National Centres. Typically the study provides data-entry software which is used in all countries. Good data-entry software provides a number of initial checks on the data that can be corrected immediately during the data-entry process. These data are then returned to the international central point where further data cleaning takes place. There are always 'extra' errors in data entry no matter how good the data-entry program. By undertaking consistency checks it is possible to identify questions on the questionnaires where an error occurred on the part of the respondent. These 'problems' are reported back to national centres; they then contact the schools for elucidation and

then send the 'correct' data back to the international data-processing centre. The necessary changes are then made. This 'cleaning' process can take a long time, especially when there are many countries in the study. However it should be mentioned that a data set from one country, where some carelessness is evident in the data collection and/or data entry, can take an inordinate time to clean. Finally, in order to account for different probabilities of selection (due to disproportionate selection across strata, inaccurate sampling frames, missing data, etc.) sampling weights have to be calculated.

All of these points should be expected to have been covered in reports on properly conducted international studies, but it is incumbent upon the reader to check this.

▒ Data analyses

Some of the analyses will be simple and others complex. If 'dummy tables' have been produced at the onset of the study, the analyses are undertaken to complete the tables. If the reader is not experienced in data analysis, it is usually wise to have experts advise him or her on the appropriateness of the analyses for the questions posed.

▒ Reports emanating from the study

The reports should be clearly written and deal with each of the policy issues in turn. The source of the data under discussion should always be clear, as should arguments concerning the interpretation of the analyses.

REFERENCES

Bishop, J. (1998) *Do curriculum-based external exit exam systems enhance student achievement?* CPRE Research Report Series RR-40, Consortium for Policy Research in Education, Graduate School of Education, University of Pennsylvania.

Goldstein, H. (1994) Educational quality and student achievement. In: P. Ribbens and E. Burridge (Eds.), *Improving education: promoting quality in schools.* Cassell: London.

Keeves, J.P.; Soydhurum, P. (1992) Specialization in science and performance at the terminal secondary school level. In: J. P. Keeves (Ed.), *The IEA study of Science III: changes in science education and achievement: 1970 to 1984.* Oxford: Pergamon Press.

Levin, H. (1998) Educational performance standards and the economy. *Educational Researcher,* 27, 4, pp. 4-10.

Postlethwaite, T.N. ; Wiley, D. . (1992) *The IEA study of Science II: science achievement in twenty-three countries.* Oxford: Pergamon Press.

Ross, K. N. (1997) Research and policy: a complex mix. *IIEP Newsletter,* XV, 1, pp. 1-4.

Wiley, D.E.; Schmidt, W.B.; Wolfe, R.E. (1992) The science curriculum and achievement. In: Postlethwaite, T.N.; Wiley, D.E. (1992) *The IEA study of Science II: science achievement in twenty-three countries.* Oxford: Pergamon Press.

GENERAL BIBLIOGRAPHY

IEA studies

Mathematics

Beaton, A.E.; Mullis, I.V.S.; Martin, M.O.; Gonzales, E.J.; Kelly, D.L.; Smith, T.A. (1996) *Mathematics achievement in the middle school years: IEA's third international mathematics and science study.* Chestnut Hill: IEA TIMSS.

Burstein, L. (ed.) (1993) *The IEA study of mathematics,* Vol. 3. Oxford: Pergamon Press.

Garden, R.A.; Robitaille, D.F. (1989) *The IEA study of mathematics II: contexts and outcomes of school mathematics.* Oxford: Pergamon Press.

Husén, T. (ed.) (1967) *International study of achievement in mathematics: a comparison of twelve countries.* Vols. 1-2. Stockholm: Almqvist and Wiksell.

Mullis, I.V.S.; Martin, M.O.; Beaton, A.E.; Gonzales, E.J.; Kelly, D.L.; Smith, T.A. (1997) *Mathematics achievement in the primary school years: IEA's third international mathematics and science study.* Chestnut Hill: IEA-TIMSS.

Travers, K.J.; Westbury, I. (1989) *The IEA study of mathematics I: international analysis of mathematics curricula.* Oxford: Pergamon Press.

Mathematics and science

Mullis, I.V.S.; Martin, M.O.; Beaton, A.E.; Gonzales, E.J.; Kelly, D.L.; Smith, T.A. (1998) ***Mathematics and science achievement in the final year of secondary school: IEA's third international mathematics and science study.*** Chestnut Hill: IEA-TIMSS.

Science

Beaton, A.E.; Martin, M.O.; Mullis, I.V.S.; Gonzales, E.J.; Smith, T.A.; Kelly, D.L. (1996) ***Science achievement in the middle school years: IEA's third international mathematics and science study.*** Chestnut Hill: IEA TIMSS.

Comber, L.C.; Keeves, J.P. (1973) ***Science education in nineteen countries: an empirical study. International Studies in Education,*** Vol. 1. New York: Wiley.

IEA (1988) ***Science achievement in seventeen countries: a preliminary report.*** Oxford: Pergamon Press.

Keeves, J.P. (1992) ***The IEA study of Science III: changes in science education and achievement: 1970 to 1984.*** Oxford: Pergamon Press.

Martin, M.O.; Mullis, I.V.S.; Beaton, A.E.; Gonzales, E.J.; Smith, T.A.; Kelly, D.A. (1997) ***Science achievement in the primary school years: the IEA's third international mathematics and science study.*** Chestnut Hill: IEA-TIMSS.

Postlethwaite, T.N. and Wiley, D.E. (1992) ***The IEA study of Science II: science achievement in twenty-three countries.*** Oxford: Pergamon Press.

Rosier, M.J.; Keeves, J.P. (1991) *The IEA study of Science I: science education and curricula in twenty-three countries.* Oxford: Pergamon Press.

Literature

Purves, A.C. (1973) *Literature education in ten countries: an empirical study. International Studies in Education,* Vol. 2. Stockholm: Almqvist and Wiksell.

Reading

Elley, W.B. (1992) *How in the world do students read?* Hamburg: IEA.

Elley, W.B. (1994) *The IEA study of reading literacy.* Oxford: Pergamon.

Thorndike, R.L. (1973) *Reading comprehension education in fifteen countries: an empirical study. International Studies in Evaluation,* Vol. 3. Stockholm: Almqvist and Wiksell.

French as a foreign language

Carroll, J.B. (1976) *The teaching of French as a foreign language in eight countries.* New York: John Wiley and Sons.

English as a foreign language

Lewis, E.G.; Massad, C. (1976) *The teaching of English as a foreign language in ten countries.* New York: John Wiley & Sons.

Civic education

Torney, J.V.; Oppenheim, A.N.; Farnen, R.F. (1976) *Civic education in ten countries: an empirical study. International Studies in Education,* Vol. 6. Stockholm: Almqvist and Wiksell.

Writing

Gorman, T.P.; Purves, A.C.; Degenhart, R.E. (eds.) (1988) ***The IEA study of written composition I: the international writing and scoring scales.*** Oxford: Pergamon Press.

Purves, A.C.; Lehmann, R.; Degenhart, R.E. (1992) ***The IEA study of written composition II: education and performance in fourteen countries.*** Oxford: Pergamon Press.

Classroom environment study

Anderson, L.W.; Ryan, D.; Shapiro, B.J. (eds.) (1989) ***The IEA classroom environment study.*** Oxford: Pergamon Press.

Computers in education

Pelgrum, W.J.; Plomp, T. (1991) ***The use of computers in education worldwide: results from the IEA computers in education survey in nineteen educational systems.*** Oxford: Pergamon Press.

Other IEA studies

Foshay, A.W. (ed.) (1962) ***Educational achievements of thirteen-year-olds in twelve countries.*** Hamburg: UNESCO Institute for Education.

Passow, A.H.; Noah, H.J.; Eckstein, M.A.; Mallea, J.R. (1976) ***The national case study: an empirical comparative study of twenty-one educational systems. International Studies in Evaluation.*** Vol. 8. Stockholm: Almqvist and Wiksell.

Peaker, G.F. (1975) ***An empirical study of education in twenty-one countries: a technical report. International Studies in Evaluation VIII.*** New York: John Wiley and Sons.

Vári, P. (ed.) (1997) *Are we similar in maths and science? A study of Grade 8 in nine Central and Eastern European countries.* Budapest: IEA.

Walker, D.A. (1976) *The IEA six-subject survey: an empirical study of education in twenty-one countries. International Studies in Evaluation,* Vol. 9. Stockholm: Almqvist and Wiksell.

SACMEQ studies

Kulpoo, D. (1998) *Mauritius: The quality of education: some policy suggestions based on a survey of schools.* SACMEQ Policy Research: Report No. 1. Paris: International Institute for Educational Planning.

Machingaidze, T.; Pfukani, P.; Shumba, S. (1998) *Zimbabwe: The quality of education: some policy suggestions based on a survey of schools.* SACMEQ Policy Research: Report No. 3. Paris: International Institute for Educational Planning.

Nassor, S.; and Mohammed, K.A. (1998) *Zanzibar: The quality of education: some policy suggestions based on a survey of schools.* SACMEQ Policy Research: Report No. 4. Paris: International Institute for Educational Planning.

Nkamba, M.; Kanyika, J. (1998) *Zambia: The quality of education: some policy suggestions based on a survey of schools.* SACMEQ Policy Research: Report No. 5. Paris: International Institute for Educational Planning.

Ross, K.N. (ed.) (1995) From educational research to educational policy: an example from Zimbabwe. Thematic issue of *International Journal of Educational Research*, Vol. 23, No. 4.

Voigts, F. (1998) *Namibia: The quality of education: some policy suggestions based on a survey of schools.* SACMEQ Policy Research: Report No. 2. Paris: International Institute for Educational Planning.

IAEP studies

Lapointe, A.E.; Mead, N.A.; Phillips, G.W. (1989) *A world of differences: an international assessment of mathematics and science.* Princeton NJ: Educational Testing Service.

Lapointe, A.E.; Askew, J.M.; Mead, N.A. (1992) *Learning science.* Princeton NJ: Educational Testing Service.

Lapointe, A.E.; Mead, N.A.; Askew, J.M. (1992) *Learning mathematics.* Princeton NJ: Educational Testing Service.

Other studies of relevance

Stevenson, H.W.; Stigler, J.W. (1992) *The learning gap: why our schools are failing and what can we learn from Japanese and Chinese education?* New York: Summit Books.

IIEP Publications and Documents

More than 1,200 titles on all aspects of educational planning have been published by the International Institute for Educational Planning. A comprehensive catalogue, giving details of their availability, includes research reports, case studies, seminar documents, training materials, occasional papers and reference books in the following subject categories:

Economics of education, costs and financing.

Manpower and employment.

Demographic studies.

Location of schools (school map) and micro-planning.

Administration and management.

Curriculum development and evaluation.

Educational technology.

Primary, secondary and higher education.

Vocational and technical education.

Non-formal, out-of-school, adult and rural education.

Disadvantaged groups.

Copies of the catalogue may be obtained from ***IIEP Publications*** on request.

57

About the International Academy of Education

The International Academy of Education is a not-for-profit scientific association that aims at fostering scholarly excellence in all fields of education by promoting educational research, its dissemination, and the implementation of its implications. Therefore, the Academy is dedicated to strengthening the contributions of research, solving critical educational problems throughout the world, and providing better communication among policy-makers, researchers and practitioners. Towards those ends, the Academy provides timely syntheses of research-based evidence of international importance, as well as critiques of research, its evidentiary basis, and its application to policy.

The seat of the Academy , which was founded in 1986, is at the Royal Academy of Science, Literature and Arts in Brussels, Belgium, and its co-ordinating centre is at Curtin University of Technology in Perth, Australia

The current members of the Board of Directors of the Academy are:

Erik De Corte,
University of Leuven, Belgium *(President)*
Barry Fraser,
Curtin University of Technology, Australia *(Executive Director)*
Jacques Hallak,
International Institute for Educational Planning, France
Michael Kirst,
Stanford University, US
Ulrich Teichler,
University of Kassel, Germany
Margaret Wang,
Temple University, US
Herbert Walberg,
University of Illinois at Chicago, US *(Vice President)*

The International Institute for Educational Planning

The International Institute for Educational Planning (IIEP) is an international centre for advanced training and research in the field of educational planning. It was established by UNESCO in 1963 and is financed by UNESCO and by voluntary contributions from Member States. In recent years the following Member States have provided voluntary contributions to the Institute: Denmark, Germany, Iceland, India, Ireland, Norway, Sweden, Switzerland and Venezuela.

The Institute's aim is to contribute to the development of education throughout the world, by expanding both knowledge and the supply of competent professionals in the field of educational planning. In this endeavour the Institute co-operates with interested training and research organizations in Member States. The Governing Board of the IIEP, which approves the Institute's programme and budget, consists of a maximum of eight elected members and four members designated by the United Nations Organization and certain of its specialized agencies and institutes.

Chairman:

Lennart Wohlgemuth (Sweden)
Director, The Nordic Africa Institute, Uppsala, Sweden.

Designated Members:

David de Ferranti
Vice President, Latin America and Caribbean region, The World Bank, Washington, USA.

Carlos Fortin
Deputy to the Secretary-General, United Nations Conference on Trade and Development (UNCTAD), Geneva, Switzerland.

Miriam J. Hirschfeld
Director, Division of Human Resources Development and Capacity Building, World Health Organization (WHO), Geneva, Switzerland.

Jeggan Senghor
Director, African Institute for Economic Development and Planning (IDEP), Dakar, Senegal.

Elected Members:

Dato'Asiah bt. Abu Samah (Malaysia)
Corporate Adviser, (Education) Lang Education, Kuala Lumpur, Malaysia.

Klaus Hüfner (Germany)
Professor, Freie Universität Berlin, Berlin, Germany.

Faïza Kefi (Tunisia)
Minister of the Environment, Ariana, Tunisia.

Tamas Kozma (Hungary)
Director-General, Hungarian Institute for Educational Research, Budapest, Hungary.

Teboho Moja (South Africa)
Visiting Professor, New York University, New York, USA.

Yolanda M. Rojas (Costa Rica)
Professor, University of Costa Rica, San José, Costa Rica.

Michel Vernières (France)
Professor, University of Paris I, Panthéon-Sorbonne, Paris, France.

Inquiries about the Institute should be addressed to:
The Office of the Director, International Institute for Educational Planning, 7-9 rue Eugène-Delacroix, 75116 Paris, France.

The book

This publication was prepared by the International Academy of Education for the IIEP, and deals with the very high-profile issue of international educational achievement studies. These studies are surrounded by complexity, misunderstanding, and questionable policy implications. This booklet examines the major questions and issues involved in international studies in a direct, concise and understandable manner.

The booklet includes the major questions that have arisen since the 1960s. All the major international studies are referenced, and examples used to demonstrate conclusions and explore trends. The authors present the evidence from several studies to answer the questions education activists and the public most often ask. Brevity is a particular virtue of this booklet in an area where technical discussions often obscure the overall trend or summary evidence.

The benefits and limitations of international achievement studies

International Academy of Education

The International Academy of Education is a not-for-profit scientific association that aims at fostering scholarly excellence in all fields of education by promoting educational research, its dissemination, and the implementation of its implications. The Academy is dedicated to strengthening the contributions of research, solving critical educational problems throughout the world, and providing better communication among policy-makers, researchers and practitioners. Towards those ends, the Academy prepares syntheses of research-based evidence of international importance, as well as critiques of research, its evidentiary basis, and its application to policy.

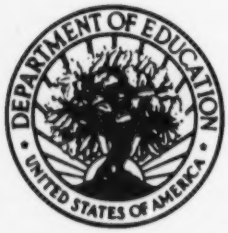
The seat of the Academy , which was founded in 1986, is at the Royal Academy of Science, Literature and Arts in Brussels, Belgium, and its co-ordinating centre is at Curtin University of Technology in Perth, Australia.

international institute
for educational planning

ISBN: 92-803-1188-3

BEST COPY AVAILABLE

EA031630



U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

ERIC

NOTICE

REPRODUCTION BASIS

☑ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").